Proof of Theorem 4
We first present two lemmas that are essential to the proof of Theorem 4, which concern the concentration of the empirical covariance matrix $\hat{C}$ around its population version $C$ and the score vector
\[
\frac{1}{n} \hat{X}^T (y - \hat{X}\beta_0) = \frac{1}{n} \hat{X}^T \eta - \frac{1}{n} \hat{X}^T (\hat{X} - X)\beta_0
\]
around zero. These lemmas can be viewed as generalizations of Lemma A.3 and inequality (A.15), respectively. For ease of presentation, we condition on the event of probability $1 - \pi_0$ that the two error bounds in Condition (C4) hold, and incorporate the probability $\pi_0$ into the result by the union bound.

Lemma S.1. Under Conditions (C4)--(C6), if $\mu_0 > 0$ and the first-stage error bounds $e_1$ and $e_2$ satisfy
\[
s(2Le_1 + e_2) \leq \frac{\alpha}{(4 - \alpha)\varphi} \land \left( \frac{\mu_0/2}{s} \right),
\]
then with probability at least $1 - \pi_0$, the following inequalities holds:
\[
\| (\hat{C}_{SS})^{-1} \|_{\infty} \leq \frac{4 - \alpha}{2(2 - \alpha)\varphi},
\]
\[
\| \hat{C}_{S\neq S}(\hat{C}_{SS})^{-1} \|_{\infty} \leq \left\{ \left( 1 - \frac{\alpha}{2} \right) \frac{\rho'(0+)}{\rho'(b_0/2)} \right\} \land (2cn^\nu),
\]
and
\[
\Lambda_{\text{min}}(\hat{C}_{SS}) > \mu\tau_0.
\]
Proof. It follows from the arguments in the proof of Lemma A.1 and Condition (C4) that
\[
\max_{1 \leq i, j \leq p} \frac{1}{n} | \hat{x}_i^T \hat{x}_j - (Z\gamma_0)^T Z\gamma_0 | \leq 2Le_1 + e_2.
\]
Consequently, by the assumption (S.1),
\[ \varphi \| \hat{C}_{SS} - C_{SS} \|_\infty \leq \varphi s(2Le_1 + e_2) \leq \frac{\alpha}{4 - \alpha} \]  
(S.5)
and
\[ \varphi \| \hat{C}_{S'S'} - C_{S'S'} \|_\infty \leq \frac{\alpha}{4 - \alpha}. \]  
(S.6)

Then inequality (S.2) follows as in the proof of Lemma A.3.

To show inequality (S.3), by (S.2), (S.5), (S.6), and Condition (C6), we have
\[
\| \hat{C}_{S'S'}(\hat{C}_{SS})^{-1} - C_{S'S'}(C_{SS})^{-1} \|_\infty 
\leq \| \hat{C}_{S'S'} - C_{S'S'} \|_\infty \| (\hat{C}_{SS})^{-1} \|_\infty + \| C_{S'S'}(C_{SS})^{-1} \|_\infty \| \hat{C}_{SS} - C_{SS} \|_\infty \| (\hat{C}_{SS})^{-1} \|_\infty 
\leq \frac{\alpha}{4 - \alpha} \frac{4 - \alpha}{2(2 - \alpha)} \varphi + \left\{ \frac{\alpha}{(1 - \alpha)} \frac{\rho'(0+)}{\rho'_\mu(b_0/2)} \right\} \wedge (cn^\nu) \frac{\alpha}{4 - \alpha} \frac{4 - \alpha}{2(2 - \alpha)} \varphi 
\leq \frac{\alpha}{2(2 - \alpha)} + \left\{ \frac{\alpha}{(1 - \alpha)} \frac{\rho'(0+)}{2(2 - \alpha) \rho'_\mu(b_0/2)} \right\} \wedge \left( \frac{c}{2} n^\nu \right) 
\leq \left\{ \frac{\alpha}{2} \frac{\rho'(0+)}{\rho'_\mu(b_0/2)} \right\} \wedge (cn^\nu),
\]
where we have used the inequalities \( \rho'(0+)/\rho'_\mu(b_0/2) \geq 1 \) and \( \alpha/(2 - \alpha) \leq 1/2 \leq cn^\nu/2. \) This, along with Condition (C6), implies (S.3).

Finally, it follows from the Hoffman–Wielandt inequality (Horn and Johnson 1985) and the assumption (S.1) that
\[ |\Lambda_{\min}(\hat{C}_{SS}) - \Lambda_{\min}(C_{SS})|^2 \leq \| \hat{C}_{SS} - C_{SS} \|_F^2 \leq s^2(2Le_1 + e_2) \leq \left( \frac{\mu_0}{2} \right)^2. \]
In view of the definition of \( \mu_0, \) inequality (S.4) follows. This completes the proof of the lemma.

**Lemma S.2.** Under Conditions (C4)–(C6), if the first-stage error bounds satisfy \( e_1 = O(1) \) and \( e_2 = O(1), \) then there exist constants \( c_0, c_1, c_2 > 0 \) such that, if we choose
\[ \mu \geq C_0n^\nu \sqrt{\log p + \log q \over n} \vee e_2, \]
where \( C_0 = c_0L \max(\sigma_{p+1}, M\sigma_{\max}, M), \) then with probability at least \( 1 - \pi_0 - c_1(pq)^{-c_2}, \) it holds that
\[ \left\| \frac{1}{n} \hat{X}^T \eta - \frac{1}{n} \hat{X}^T (\hat{X} - X) \beta_0 \right\|_\infty \leq \frac{\alpha}{6cn^\nu} \mu \rho'(0+). \]  
(S.7)

**Proof.** As in the proof of Lemma A.2, we write \( n^{-1} \hat{X}^T (\hat{X} - X) \eta = T_1 + \cdots + T_6. \) Letting \( t_0 = \alpha \mu \rho'(0+)/6cn^\nu, \) we bound the six terms similarly as follows:
\[ P \left( \| T_1 \|_\infty \geq \frac{t_0}{6} \right) \leq P \left( \left\| \frac{1}{n} Z^T \eta \right\|_\infty \geq \frac{t_0}{6e_1} \right) \leq q \exp \left\{ - \frac{n}{2\sigma_{p+1}^2} \left( \frac{t_0}{6e_1} \right)^2 \right\}. \]
Define the function $f$ and $\mu \rho$ and then show that thus obtained $\hat{\theta}$ of Theorem 3, (S.8) with the inequalities in Lemmas S.1 and S.2 hold. Using similar arguments to those in the proof of Theorem 4, it suffices to find a $(S.8)$–(S.10) hold. Let $\hat{\theta}$ and $K$ exist constants $c_0, c_1, c_2 > 0$ such that, if we choose

$$
\mu \geq C_0 n^\nu \sqrt{\frac{\log p + \log q}{n}} \vee e_2,
$$

where $C_0 = c_0 L \max(\sigma_{p+1}, M\sigma_{\max}, M)$, then with probability at least $1 - \pi_0 - c_1(pq)^{-c_2}$, the desired inequality holds. The completes the proof of the lemma.

**Proof of Theorem 4.** One can easily show that $\hat{\beta} \in \mathbb{R}^p$ is a strict local minimizer of problem (4) if the following conditions hold:

$$
\frac{1}{n} \hat{X}_S^T (y - \hat{X}\hat{\beta}) = \mu \rho'_\mu(|\hat{\theta}_S|) \circ \text{sgn}(\hat{\theta}_S),
$$

(S.8)

and

$$
\left\| \frac{1}{n} \hat{X}_{\hat{S}}^T (y - \hat{X}\hat{\beta}) \right\|_\infty < \mu \rho'(0+),
$$

(S.9)

and

$$
\Lambda_{\min}(\hat{C}_{\hat{S}\hat{S}}) > \mu \tau(\rho_\mu; \hat{\theta}_S),
$$

(S.10)

where $\circ$ denotes the Hadamard (entrywise) product, and $| \cdot |$, $\rho'_\mu(\cdot)$, and $\text{sgn}(\cdot)$ are applied componentwise. It suffices to find a $\hat{\beta} \in \mathbb{R}^p$ with the desired properties such that conditions (S.8)–(S.10) hold. Let $\hat{\beta}_{\hat{S}} = 0$. The idea of the proof is to first determine $\hat{\beta}_S$ from (S.8), and then show that thus obtained $\hat{\beta}$ also satisfies (S.9) and (S.10).

From now on, we condition on the event of probability at least $1 - \pi_0 - c_1(pq)^{-c_2}$ that the inequalities in Lemmas S.1 and S.2 hold. Using similar arguments to those in the proof of Theorem 3, (S.8) with $\hat{S}$ replaced by $S$ can be written in the form

$$
\hat{\beta}_S - \beta_{0S} = (\hat{C}_{SS})^{-1} \left\{ \frac{1}{n} \hat{X}_S^T \eta - \frac{1}{n} \hat{X}_S^T (\hat{X}_S - X_S) \beta_{0S} - \mu \rho'_\mu(|\hat{\theta}_S|) \circ \text{sgn}(\hat{\theta}_S) \right\}. 
$$

(S.11)

Define the function $f : \mathbb{R}^s \to \mathbb{R}^s$ by $f(\theta) = \beta_{0S} + (\hat{C}_{SS})^{-1} \{n^{-1} \hat{X}_S^T \eta - n^{-1} \hat{X}_S^T (\hat{X}_S - X_S) \beta_{0S} - \mu \rho'_\mu(|\theta|) \circ \text{sgn}(\theta) \}$, and let $\mathcal{K}$ denote the hypercube $\{ \theta \in \mathbb{R}^s : \| \theta - \beta_{0S} \|_\infty \leq 7 \varphi \mu \rho'(0+)/4 \}$. 

3
It follows from (S.2), (S.7), and Condition (C4) that, for $\theta \in \mathcal{K}$,
\[
\|f(\theta) - \beta_{0S}\|_\infty \leq \|(\mathcal{C}_{SS})^{-1}\|_{\infty} \left\{ \left\| \frac{1}{n} \hat{X}_S^T \eta - \frac{1}{n} \hat{X}_S^T (\hat{X}_S - X_S) \beta_{0S} \right\|_\infty + \mu \rho'(0+) \right\}
\leq \frac{4 - \alpha}{2(2 - \alpha)^\varphi} \left\{ \frac{\alpha}{6cn^\nu} \mu \rho'(0+) + \mu \rho'(0+) \right\}
\leq \frac{3}{2} \varphi \left\{ \frac{1}{6} \mu \rho'(0+) + \mu \rho'(0+) \right\} = \frac{7}{4} \varphi \mu \rho'(0+),
\]
that is, $f(\mathcal{K}) \subseteq \mathcal{K}$. Also, the last inequality and the assumption (14) imply that for $\theta \in \mathcal{K}$, $\|\theta - \beta_{0S}\|_\infty \leq b_0/2$, and hence $\text{sgn}(\theta) = \text{sgn}(\beta_{0S})$. Thus, in view of Condition (C4), $f$ is a continuous function on the convex, compact hypercube $\mathcal{K}$. An application of Brouwer’s fixed point theorem yields that equation (S.11) has a solution $\hat{\beta}_S$ in $\mathcal{K}$. Moreover, $\text{sgn}(\hat{\beta}_S) = \text{sgn}(\beta_{0S})$, so that $\hat{\beta} = \hat{\beta}_S$. Therefore, we have found a $\hat{\beta}$ that satisfies the desired properties and (S.8).

To verify that $\hat{\beta}$ satisfies (S.9), by substituting (S.11), we write
\[
\frac{1}{n} \hat{X}_S^T (y - \hat{X}\beta) = \frac{1}{n} \hat{X}_S^T \eta - \frac{1}{n} \hat{X}_S^T (\hat{X}_S - X_S) \beta_{0S} - \hat{C}_{SS}(\mathcal{C}_{SS})^{-1} \left\{ \frac{1}{n} \hat{X}_S^T \eta - \frac{1}{n} \hat{X}_S^T (\hat{X}_S - X_S) \beta_{0S} - \mu \rho'_\mu(\hat{\beta}_S) \circ \text{sgn}(\hat{\beta}_S) \right\}.
\]
Also, we have $\|\hat{\beta}_S\|_\infty = \|\hat{\beta}_{0S} + (\hat{\beta}_S - \beta_{0S})\|_\infty \geq \|\hat{\beta}_{0S}\|_\infty - \|\hat{\beta}_S - \beta_{0S}\|_\infty \geq b_0 - b_0/2 = b_0/2$. This, together with (S.3), (S.7), and Condition (C4), leads to
\[
\left\| \frac{1}{n} \hat{X}_S^T (y - \hat{X}\beta) \right\|_\infty \leq \left\| \frac{1}{n} \hat{X}_S^T \eta - \frac{1}{n} \hat{X}_S^T (\hat{X}_S - X_S) \beta_{0S} \right\|_\infty + \|\hat{C}_{SS}(\mathcal{C}_{SS})^{-1}\|_{\infty}
\times \left\{ \left\| \frac{1}{n} \hat{X}_S^T \eta - \frac{1}{n} \hat{X}_S^T (\hat{X}_S - X_S) \beta_{0S} \right\|_\infty + \mu \rho'_\mu(b_0/2) \right\}
\leq \frac{\alpha}{6cn^\nu} \mu \rho'(0+) + 2cn^\nu \cdot \frac{\alpha}{6cn^\nu} \mu \rho'(0+) + \left(1 - \frac{\alpha}{2}\right) \frac{\rho'(0+)}{\rho'_\mu(b_0/2)} \cdot \mu \rho'_\mu(b_0/2)
\leq \frac{\alpha}{6} \mu \rho'(0+) + \frac{\alpha}{3} \mu \rho'(0+) + \left(1 - \frac{\alpha}{2}\right) \mu \rho'(0+).
\]
Finally, it follows from (S.4) and the definition of $\tau_0$ that $\Lambda_{min}(\hat{\mathcal{C}}_{SS}) > \mu \tau_0 \geq \mu \tau(\rho'_\mu; \hat{\beta}_S)$, which verifies (S.10) and completes the proof.